

A Cheeger Inequality for the Graph Connection Laplacian

Afonso S. Bandeira*

Amit Singer†

Daniel A. Spielman‡

March 2, 2013

Abstract

The $O(d)$ Synchronization problem consists of estimating a set of unknown orthogonal transformations O_i from noisy measurements of a subset of the pairwise ratios $O_i O_j^{-1}$. We formulate and prove a Cheeger-type inequality that relates a measure of how well it is possible to solve the $O(d)$ synchronization problem with the spectra of an operator, the graph Connection Laplacian. We also show how this inequality provides a worst case performance guarantee for a spectral method to solve this problem.

Keywords: $O(d)$ Synchronization, Cheeger Inequality, Graph Connection Laplacian, Vector Diffusion Maps.

1 Introduction

The Synchronization problem over a group \mathcal{G} on a graph $G = (V, E)$ consists of estimating a group potential $g : V \rightarrow \mathcal{G}$ on the vertices given, for each edge $(i, j) \in E$, a measurement of the offset $\rho_{ij} = g_i g_j^{-1}$. If the measurements are noiseless and the graph is connected then the problem can be easily solved by considering a spanning tree in G and sequentially determining the group potential values by traveling along the tree and sequentially multiplying the offsets. The potential obtained is uniquely determined up to an overall group translation, that corresponds to fixing its value on one node.

The case $\mathcal{G} = O(d)$, the group of $d \times d$ orthogonal matrices (rotations, reflections and compositions of both on \mathbb{R}^d), is of particular interest in several applications. For example, when $d = 1$, i.e. $\mathcal{G} = O(1) \cong \mathbb{Z}/2\mathbb{Z}$, the synchronization solution can be used to determine whether a manifold is orientable [14]. An interesting aspect of the $O(1)$ synchronization problem is that it is a generalization of the max-cut problem [17] as we will see later. The $O(d)$ synchronization also plays a major role in a certain algorithm for the sensor network localization problem [7]. The, very similar, synchronization in $SO(d)^1$ also has several applications; for $d = 3$, it can be used, e. g., for global alignment of 3-D scans [18] and the $SO(2)$ problem plays a major role in new developments on the Cryo-electron microscopy problem (see [12, 15]). Several more applications for $SO(2)$ synchronization are presented in [11] and references therein.

When there is noise in the offset measurements, the spanning tree method will suffer from accumulation of errors. Thus, there is the need for methods that use the redundancy in the information given and globally estimate all values simultaneously. Such a method is proposed and analyzed in [11] for $\mathcal{G} = SO(2)$. This method can be adapted to deal with other groups of orthogonal matrices, as we will see later. In a nutshell,

*Program in Applied and Computational Mathematics (PACM), Princeton University, Princeton, NJ 08544, USA (ajs@math.princeton.edu).

†PACM and Department of Mathematics, Princeton University, Princeton, NJ 08544, USA (amits@math.princeton.edu).

‡Department of Applied Mathematics and Department of Computer Science, Yale University, PO Box 208285, New Haven, CT 06520-8285, USA (spielman@cs.yale.edu).

¹The group of rotations in \mathbb{R}^d .

the method consists of constructing an operator (a matrix) whose eigenvectors can be used to estimate the group potential. In [11], the resilience to a certain type of random noise is analyzed. There is, however, no worst case guarantee.

Nevertheless, some other spectral methods have worst case performance guarantees. In fact, Cheeger's Inequality for graphs [1, 2] provides such a guarantee for spectral clustering. It relates the spectra of the graph Laplacian with bottlenecks in the graph. Informally, it says that clusters are the only obstacles to the rapid mixing of a random walk on a graph. In the present paper we provide a similar guarantee for $O(d)$ Synchronization via a Cheeger-like inequality for the graph Connection Laplacian.

Recently, a notion of vector diffusion on graphs has been introduced [13]. One of its purposes is to deal with data for which there is a natural rotation group action². Very roughly, the diffusion process, instead of moving point masses between vertices, moves vectors and rotates them depending on the relative rotation between the data points. If there is a cycle with an inconsistent set of relative rotations then cancellations will take place in the diffusion process and there will be loss of mass. This loss of mass can be measured by the spectra of the graph Connection Laplacian [13]. In the present paper we show, for $\mathcal{G} = O(d)$, a type of a Cheeger Inequality (Theorem 10) relating the spectra of this operator with a measure of inconsistency on the offsets over cycles of the graph, showing that they are, in fact, responsible for the mass loss of the vector diffusion process.

The inequality shown in the present paper provides a worst case performance guarantee for a spectral method (Algorithm 16) to solve the $O(d)$ synchronization problem. It is worth noting that, in a different setup, Trevisan [17] showed a particular case of this inequality, when the group is $O(1) \cong \mathbb{Z}/2\mathbb{Z}$ and all the offsets are -1 . In that case the problem is equivalent to the Max-Cut problem. Therefore, this $O(1)$ inequality gives a performance guarantee for a spectral method to solve Max-Cut [17].

Throughout the paper we use the notation $[n]$ to refer to $\{1, \dots, n\}$. Also, we make use of several matrix and vector notations. Given a matrix A we denote by $\|A\|_F$ its Frobenius norm. If A is symmetric we denote by $\lambda_1(A), \lambda_2(A), \dots$ its eigenvalues in increasing order. Assuming further that A is positive semi-definite we define the A -inner product of vectors x and y as $\langle x, y \rangle_A = x^T A y$ (and say that two vectors are A -orthogonal if it is zero). Also, we define the A -norm of x as $\|x\|_A = \sqrt{\langle x, x \rangle_A}$. Given $x \in (\mathbb{R}^d)^n$ we denote by x_i the i -th $d \times 1$ block of x (that will correspond to the value of x on the vertex i). Also, for $t \geq 0$, we define x^t as the $(\mathbb{R}^d)^n$ vector such that,

$$x_i^t = \begin{cases} 0 & \text{if } \|x_i\|^2 < t \\ \frac{x_i}{\|x_i\|} & \text{if } \|x_i\|^2 \geq t. \end{cases}$$

Finally, \tilde{x} denotes x^0 , with the small nuance that if $x_i = 0$ for some i then \tilde{x}_i can simply be any unit vector in \mathbb{R}^d .

The rest of this section includes mathematical preliminaries that are needed in later sections. Section 2 consists of the solution to a subproblem and Section 3 contains both the formulation and proof of our main result. We discuss an Algorithm for $O(d)$ synchronization in Section 4 and show a few tightness results in Section 5. We end with some concluding remarks and a few open problems in Section 6.

1.1 Preliminaries

Let $G = (V, E)$ be an undirected weighted graph with n vertices. Let W_0 be the weighted adjacency matrix of G and D_0 be the degree matrix, a diagonal matrix with diagonal elements $(D_0)_{ii} = \sum_{j=1}^n w_{ij}$. The matrix $D_0^{-1}W_0$ is the transition probability matrix of the random walk in G , whose transition probabilities are proportional to the edge weights. The graph Laplacian of G is defined as $L_0 = D_0 - W_0$. It is also useful to consider the normalized graph Laplacian $\mathcal{L}_0 = D_0^{-1/2}L_0D_0^{-1/2} = I - D_0^{-1/2}W_0D_0^{-1/2}$. Both L_0 and \mathcal{L}_0 are

²For example, a data set of hand written digits where one wants to determine whether two written digits are the same up to rotation.

symmetric positive semi-definite matrices. The eigenvalues of \mathcal{L}_0 encode important information about the random walk. In fact, the second smallest eigenvalue³ is a good measure of how well the random walk mixes. More specifically, the smaller $\lambda_2(\mathcal{L}_0)$, the slower the convergence to the limiting stationary distribution.

It is clear that clusters, i.e. sets that are very internally connected but poorly connected to the rest of the graph, will constitute obstacles to rapid mixing of the random walk, since the probability mass might be trapped inside such a set for a while. Cheeger's Inequality shows that, in some sense, these sets are the only obstacles to rapid mixing (several different proofs for this inequality can be found in [5]).

Theorem 1 ([1, 2]). *Let $G = (V, E)$ be a graph and \mathcal{L}_0 its normalized graph Laplacian. Then*

$$\frac{1}{2}\lambda_2(\mathcal{L}_0) \leq \min_{S \subset V} \frac{\text{cut}(S)}{\min\{\text{vol}(S), \text{vol}(S^c)\}} \leq \sqrt{2\lambda_2(\mathcal{L}_0)},$$

where $\text{cut}(S) = \sum_{i \in S, j \notin S} w_{ij}$ and $\text{vol}(S) = \sum_{i \in S} d_i$, with $d_i = \sum_{j \in V} w_{ij}$, the degree of the i -th vertex.

If, associated to each edge $(i, j) \in E$, there is an orthogonal transformation⁴ $\rho_{ij} \in O(d)$, one might be interested in considering a random walk that takes the transformations into account. One way of doing this is by defining a random walk that, instead of moving point masses, moves a vector from vertex to vertex and transforms it via the orthogonal transformation associated with the edge. A Laplacian for this diffusion, the graph connection Laplacian, was introduced in [13]. The construction goes as follows: define the symmetric matrix $W_1 \in \mathbb{R}^{dn \times dn}$, such that the (i, j) -th $d \times d$ block is given by $(W_1)_{ij} = w_{ij}\rho_{ij}$, where w_{ij} is the weight of the edge (i, j) . Also, let $D_1 \in \mathbb{R}^{dn \times dn}$, be a diagonal matrix such that $(D_1)_{ii} = d_i I_{d \times d}$. The graph Connection Laplacian L_1 is defined as $L_1 = D_1 - W_1$, and the normalized graph Connection Laplacian as

$$\mathcal{L}_1 = I - D_1^{-1/2} W_1 D_1^{-1/2}.$$

One natural motivation for this random walk is the setting in which the vertices of the graph represent a data set $Z = \{z_1, \dots, z_n\}$ with an action of $O(d)$ defined on it. Such an example is a set of images and their rotations and reflections (corresponding to $d = 2$). On this type of data set, one is interested in comparing two data points when they are optimally aligned. Thus, we set ρ_{ij} to be the element in $O(d)$ that makes z_j as close as possible (in some meaningful distance) to z_i , meaning

$$\rho_{ij} = \underset{\rho \in O(d)}{\text{argmin}} \|z_i - \rho z_j\|,$$

and set the affinity between the data points accordingly.

In this case, the vector random walk (or diffusion) can be thought of as a random walk on the orbits of the group action. This kind of walk might be more meaningful because it attempts to quotient out the group action. If all the data points lie in a single orbit then there exists a group potential $O : V \rightarrow O(d)$ such that, $\forall (i, j) \in E$, $\rho_{ij} = O_i O_j^{-1}$. Finding such a group potential is the objective of the Synchronization problem. In this case the walk should have a stationary distribution, corresponding to the null space of the graph Connection Laplacian. In fact, by computing the Rayleigh quotient

$$\frac{x^T L_1 x}{x^T D_1 x} = \frac{1}{2} \frac{\sum_{(i,j) \in E} w_{ij} (\|x_i\|^2 + \|x_j\|^2 - 2x_i^T \rho_{ij} x_j)}{\sum_{i \in V} d_i} = \frac{1}{2} \frac{\sum_{(i,j) \in E} w_{ij} \|x_i - \rho_{ij} x_j\|^2}{\sum_{i \in V} d_i}, \quad (1)$$

one notes that each column of the $dn \times d$ matrix representation of the potential O is an eigenvector (in the null space) of the graph Connection Laplacian. This observation motivates the spectral method for Synchronization. As it turns out, this method tends to be very robust to outliers [11], meaning that even if several measurements are extremely corrupted the potential recovered is likely to still be very close to the original one.

³We note that the smallest eigenvalue is always 0.

⁴Recall that $O(d)$ is the group of orthogonal transformations in \mathbb{R}^d .

In fact, we will be interested in cases for which a rotation group potential may not exist (due to, e.g., errors in measurements) and the objective will be to estimate the “best possible” potential. Inspired by (1) we define a penalty function for the synchronization problem as, given a potential O ,

$$\nu(O) = \frac{1}{2d} \frac{1}{\text{vol}(G)} \sum_{ij} w_{ij} \|O_i - \rho_{ij} O_j\|_F^2.$$

As $w_{ij} = 0$ when $(i, j) \notin E$, we can sum over all pairs of vertices without loss of generality. Since our objective is to minimize this function, we define the $O(d)$ frustration ℓ_2 constant of G as

$$\nu_G = \min_{O: V \rightarrow O(d)} \nu(O).$$

It is unclear which penalty function one should consider. Considering the sum of the squares of the Frobenius norms of the incompatibilities (the ℓ_2 penalty function) is more close in spirit to (1) and thus more similar to what the spectral method will try to minimize, which suggests that this penalty function is easier to analyze. On the other hand, considering the sum of the Frobenius norms of the incompatibilities (in some sense, an ℓ_1 penalty function) will induce sparsity, meaning that it will favor candidate solutions for which some edges are perfectly correct even if it has some edges with a large error. This might make more sense in some noise models on the edge measurements. In fact, the noise model analyzed in [11] consists of a few randomly chosen edges having a measurement that is randomly drawn with respect to the uniform distribution in the space of possible measurements (in our case $O(d)$, in [11] $SO(2)$), therefore the original rotation potential will perfectly agree with some edges and have a large error on others. This suggests that considering the ℓ_1 penalty function may be more suitable to recover the original rotation potential. For that reason we also analyze it, where we formally define the ℓ_1 penalty function as

$$\vartheta(O) = \frac{1}{\sqrt{d} \text{vol}(G)} \sum_{ij} w_{ij} \|O_i - \rho_{ij} O_j\|_F,$$

and the $O(d)$ frustration ℓ_1 constant of G as

$$\vartheta_G = \min_{O: V \rightarrow O(d)} \vartheta(O).$$

2 The \mathbb{S}^{d-1} localization problem

Let us consider a simpler version of the problem. Suppose we just want to estimate one column of the synchronization solution. We can state it as estimating a function $v : V \rightarrow \mathbb{S}^{d-1}$ such that, for $(i, j) \in E$, $v_i = \rho_{ij} v_j$. This can be thought of as a localization problem on \mathbb{S}^{d-1} , the unit sphere in \mathbb{R}^d .

Again, such a function may not exist in general. Similarly to the penalty function for $O(d)$ synchronization we define an ℓ_1 penalty function $\zeta(v)$ and an ℓ_2 version $\eta(v)$ as, respectively,

$$\zeta(v) = \frac{1}{2} \frac{\sum_{ij} w_{ij} \|v_i - \rho_{ij} v_j\|}{\sum_i d_i \|v_i\|} \quad \text{and} \quad \eta(v) = \frac{1}{2} \frac{\sum_{ij} w_{ij} \|v_i - \rho_{ij} v_j\|^2}{\sum_i d_i \|v_i\|^2}.$$

We then define ζ_G , the \mathbb{S}^{d-1} frustration ℓ_1 constant of G , as

$$\zeta_G = \min_{v: V \rightarrow \mathbb{S}^{d-1}} \zeta(v),$$

and similarly, we define η_G , the \mathbb{S}^{d-1} frustration ℓ_2 constant of G , as

$$\eta_G = \min_{v: V \rightarrow \mathbb{S}^{d-1}} \eta(v).$$

The spectral approach to this problem is motivated by the fact that $\lambda_1(\mathcal{L}_1) = \min_{x \in \mathbb{R}^{dn}} \frac{x^T L_1 x}{x^T D_1 x} = \min_{x: V \rightarrow \mathbb{R}^d} \eta(x)$.

2.1 The partial frustration constants

Our objective is to relate $\lambda_1(\mathcal{L}_1)$ with ζ_G and η_G . Consider, however, the following example: a graph consisting of two disjoint components, one whose ρ_{ij} measurements are perfectly compatible and another one on which they are not. Its graph Connection Laplacian would have a non-zero vector in its null space, corresponding to just synchronizing the compatible component and being zero on the other (thus $\lambda_1(\mathcal{L}_1) = 0$). On the other hand, the constraint that v has to take values on \mathbb{S}^{d-1} , will enforce it to try to synchronize the incompatible part and will cause ζ_G and η_G to be bounded away from zero. This example suggests a relaxation of the frustration constant to allow vertices to not be labeled (labeled with 0). We thus define ζ_G^* , the partial \mathbb{S}^{d-1} frustration ℓ_1 constant of G , as

$$\zeta_G^* = \min_{v: V \rightarrow \mathbb{S}^{d-1} \cup \{0\}} \zeta(v), \quad (2)$$

and its ℓ_2 counterpart as $\eta_G^* = \min_{v: V \rightarrow \mathbb{S}^{d-1} \cup \{0\}} \eta(v)$.

We will show the following.

Theorem 2. *Let $G = (V, E)$ be a graph. Given a function $\rho : E \rightarrow O(d)$, let ζ_G^* and η_G^* be, respectively, the partial \mathbb{S}^{d-1} frustration ℓ_1 and ℓ_2 constants of G and $\lambda_1(\mathcal{L}_1)$ the smallest eigenvalue of the normalized graph Connection Laplacian. Then*

$$\frac{1}{2}\lambda_1(\mathcal{L}_1) \leq \zeta_G^* \leq \sqrt{\frac{5}{2}\lambda_1(\mathcal{L}_1)} \quad \text{and} \quad \lambda_1(\mathcal{L}_1) \leq \eta_G^* \leq 2\sqrt{\frac{5}{2}\lambda_1(\mathcal{L}_1)}, \quad (3)$$

Furthermore, if $d = 1$, the stronger inequalities hold, $\zeta_G^* \leq \sqrt{2\lambda_1(\mathcal{L}_1)}$ and $\eta_G^* \leq 2\sqrt{2\lambda_1(\mathcal{L}_1)}$.

An $O(1)$ version of this inequality was shown in [17], when ρ is the constant function equal to -1 , in the context of the Max-cut problem.

It is trivial to check that, for any $x \in \mathbb{R}^{dn}$, and $t \geq 0$, $\frac{1}{2}\eta(x^t) \leq \zeta(x^t)$. This observation, together with the fact that the eigenvalue problem is a relaxation of η_G^* , shows both left inequalities and the fact that the first right inequality implies the second in (3). The first right inequality is a direct application of the following Lemma which is obtained by an adaptation of an argument in [17].

Lemma 3. *Given $x \in \mathbb{R}^{dn}$ there exists $t > 0$ such that*

$$\zeta(x^t) \leq \sqrt{\frac{5}{2}\eta(x)}.$$

Moreover, if $d = 1$ the right-hand side can be replaced by $\sqrt{2\eta(x)}$.

In order to prove this Lemma we need the following inequality.

Proposition 4. *For any y and z unit vectors in \mathbb{R}^d , the following holds for any $\alpha \geq 1$,*

$$\|y - z\| + \alpha^2 - 1 \leq \frac{\sqrt{5}}{2}\|y - \alpha z\|(1 + \alpha).$$

Proof. Let $t = \|y - z\|$, which implies that $0 \leq t \leq 2$. Since y and z are unit vectors it is straightforward to check that $\|y - \alpha z\| = \sqrt{1 + \alpha^2 - 2\alpha(1 - \frac{1}{2}t^2)}$. Thus, it suffices to show

$$t + \alpha^2 - 1 \leq \frac{\sqrt{5}}{2}\sqrt{1 + \alpha^2 - 2\alpha\left(1 - \frac{1}{2}t^2\right)}(1 + \alpha), \quad (4)$$

for all $0 \leq t \leq 2$ and $\alpha \geq 1$. Since both sides of (4) are positive, it is enough to show the inequality with both sides squared. Squaring and rearranging yields,

$$t^2 + 2t\alpha^2 - 2t + (\alpha^2 - 1)^2 \leq \frac{5}{4} ((\alpha^2 - 1)^2 + \alpha t^2 (1 + \alpha)^2).$$

This is equivalent to the non-negativity, in the interval $[0, 2]$, of a certain quadratic function of t :

$$\left(\frac{5}{4} \alpha (1 + \alpha)^2 - 1 \right) t^2 - (2\alpha^2 - 2) t + \frac{1}{4} (\alpha^2 - 1)^2 \geq 0.$$

Since, for $\alpha \geq 1$,

$$(2\alpha^2 - 2)^2 - 4 \left(\frac{5}{4} \alpha (1 + \alpha)^2 - 1 \right) \frac{1}{4} (\alpha^2 - 1)^2 = (\alpha^2 - 1)^2 \left(4 - \frac{5}{4} \alpha (1 + \alpha)^2 + 1 \right) \leq 0,$$

the quadratic is always non-negative and thus non-negative in $[0, 2]$. \square

Proof. [of Lemma 3] Let us suppose, without loss of generality, that x is normalized so that $\max_i \|x_i\| = 1$. We will use a probabilistic argument. Let us consider the random variable t drawn uniformly from $[0, 1]$ and recall that x^t is defined by $x_i^t = \frac{x_i}{\|x_i\|}$ if $\|x_i\|^2 > t$ or $x_i^t = 0$ if $\|x_i\|^2 \leq t$. We will show that $\frac{1}{2} \frac{\mathbb{E} \sum_{ij} w_{ij} \|x_i^t - \rho_{ij} x_j^t\|}{\mathbb{E} \sum_i d_i \|x_i^t\|} \leq \sqrt{\frac{5}{2} \eta(x)}$, which implies that at least one of the realizations of t must satisfy the inequality, and proves the Lemma.

We start by showing that, for each edge (i, j) ,

$$\mathbb{E} \|x_i^t - \rho_{ij} x_j^t\| \leq \frac{\sqrt{5}}{2} \|x_i - \rho_{ij} x_j\| (\|x_i\| + \|x_j\|). \quad (5)$$

Without loss of generality we can consider $\rho_{ij} = I$ and $\|x_j\| \leq \|x_i\|$ and get,

$$\mathbb{E} \|x_i^t - x_j^t\| = \|x_j\|^2 \left\| \frac{x_i}{\|x_i\|} - \frac{x_j}{\|x_j\|} \right\| + (\|x_i\|^2 - \|x_j\|^2).$$

Thus, it suffices to show

$$\|x_j\|^2 \left\| \frac{x_i}{\|x_i\|} - \frac{x_j}{\|x_j\|} \right\| + (\|x_i\|^2 - \|x_j\|^2) \leq \frac{\sqrt{5}}{2} \|x_i - x_j\| (\|x_i\| + \|x_j\|),$$

which is a consequence of Proposition 4 for $y = \frac{x_j}{\|x_j\|}$, $z = \frac{x_i}{\|x_i\|}$ and $\alpha = \frac{\|x_i\|}{\|x_j\|}$. Now, using (5), the linearity of expectation, and the Cauchy-Schwartz inequality we have

$$\begin{aligned} \mathbb{E} \sum_{ij} w_{ij} \|x_i^t - \rho_{ij} x_j^t\| &\leq \frac{\sqrt{5}}{2} \sum_{ij} w_{ij} \|x_i - \rho_{ij} x_j\| (\|x_i\| + \|x_j\|) \\ &\leq \frac{\sqrt{5}}{2} \sqrt{\sum_{ij} w_{ij} \|x_i - \rho_{ij} x_j\|^2} \sqrt{\sum_{ij} w_{ij} (\|x_i\| + \|x_j\|)^2}. \end{aligned}$$

Since $\sum_{ij} w_{ij} \|x_i - \rho_{ij} x_j\|^2 = 2\eta(x) \sum_i d_i \|x_i\|^2$ and

$$\sum_{ij} w_{ij} (\|x_i\| + \|x_j\|)^2 \leq 2 \sum_{ij} w_{ij} (\|x_i\|^2 + \|x_j\|^2) = 4 \sum_i d_i \|x_i\|^2,$$

we have

$$\mathbb{E} \sum_{ij} w_{ij} \|x_i^t - \rho_{ij} x_j^t\| \leq \frac{\sqrt{5}}{2} \sqrt{8\eta(x)} \sum_i d_i \|x_i\|^2 = \frac{\sqrt{5}}{2} \sqrt{8\eta(x)} \mathbb{E} \sum_i d_i \|x_i^t\| = 2\sqrt{\frac{5}{2} \eta(x)} \mathbb{E} \sum_i d_i \|x_i^t\|,$$

which completes the proof. When $d = 1$ the sharper result can be obtained by noting that (5) holds even without the $\frac{\sqrt{5}}{2}$ factor. \square

It is worth noting that the proof of Lemma 3 provides an algorithm for finding $v : V \rightarrow \mathbb{S}^{d-1} \cup \{0\}$ such that $\zeta(v) \leq \sqrt{\frac{5}{2}\lambda_1(\mathcal{L}_1)} \leq \sqrt{\frac{5}{2}\zeta_G^*}$. It consists of setting x as the minimizer of $\eta(x)$ and finding t such that $\zeta(x^t) \leq \sqrt{\frac{5}{2}\lambda_1(\mathcal{L}_1)} \leq \sqrt{\frac{5}{2}\zeta_G^*}$. This algorithm is computationally tractable because minimizing $\eta(\cdot)$ is equivalent to solving an eigenvalue problem and, instead of testing x^t for every $t \geq 0$, one can limit it to the values t for which there exists $i \in [n]$ such that $\|x_i\| = t$. Trevisan [17] iteratively performs this procedure, in the subgraph composed of the vertices left unlabeled by the previous iteration, in order to label the entire graph.

2.2 The frustration constants for localization in \mathbb{S}^{d-1}

In this section we obtain bounds for ζ_G and η_G . The intuition given to justify the relaxation to partial frustration constants was based on the possible poor connectivity of the graph (small spectral gap). In this section we show that poor connectivity, as measured by a small spectral gap in the normalized graph Laplacian, is the only condition under which one can have large discrepancy between the frustration constants and the spectra of the graph Connection Laplacian. We will show that, as long as the spectral gap is bounded away from zero, one can in fact control the full frustration constants. The formal result is presented in the following Theorem.

Theorem 5. *Let $G = (V, E)$ be a graph. Given a function $\rho : E \rightarrow O(d)$, let ζ_G and η_G be, respectively, the \mathbb{S}^{d-1} frustration ℓ_1 and ℓ_2 constants of G , $\lambda_1(\mathcal{L}_1)$ the smallest eigenvalue of the normalized graph Connection Laplacian and $\lambda_2(\mathcal{L}_0)$ the second smallest eigenvalue of the normalized graph Laplacian. Then,*

$$\lambda_1(\mathcal{L}_1) \leq \eta_G \leq 44 \frac{\lambda_1(\mathcal{L}_1)}{\lambda_2(\mathcal{L}_0)},$$

$$\text{and } \frac{1}{2}\lambda_1(\mathcal{L}_1) \leq \zeta_G \leq 11\sqrt{\frac{\lambda_1(\mathcal{L}_1)}{\lambda_2(\mathcal{L}_0)}}.$$

We start by stating and proving a few results that are going to be important to both the proof of Theorem 5 and the forthcoming sections.

Lemma 6. *Given $x \in \mathbb{R}^{dn}$, there exists $\alpha_x \geq 0$, such that $r_x = x - \alpha_x \tilde{x}$ satisfies $\|r_x\|_{D_1}^2 \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \|x\|_{D_1}^2$.*

Proof. Let us define $n_x \in \mathbb{R}^n$ by $(n_x)_i = \|x_i\|$. We now set $\alpha_x = \arg\min_{\alpha} \|n_x - \alpha \mathbf{1}\|_{D_0}$. A simple calculation reveals that this gives

$$\alpha_x = \frac{\mathbf{1}^T D_0 n_x}{\mathbf{1}^T D_0 \mathbf{1}}.$$

Since n_x is a non-negative vector, α_x is non-negative as well. Let us also define $u_x \in \mathbb{R}^n$ so that $(r_x)_i = (u_x)_i \tilde{x}_i$. This implies that $u_x = n_x - \frac{\mathbf{1}^T D_0 n_x}{\mathbf{1}^T D_0 \mathbf{1}} \mathbf{1}$. Thus,

$$u_x^T L_0 u_x = n_x^T L_0 n_x = \frac{1}{2} \sum_{ij} w_{ij} (\|x_i\| - \|x_j\|)^2 \leq \frac{1}{2} \sum_{ij} w_{ij} \|x_i - \rho_{ij} x_j\|^2 = \eta(x) \|x\|_{D_1}^2$$

Since $u_x^T D_0 \mathbf{1} = 0$, we have $\frac{(u_x)^T L_0 u_x}{\|u_x\|_{D_0}^2} \geq \lambda_2(\mathcal{L}_0)$. This shows that $\|r_x\|_{D_1}^2 = \|u_x\|_{D_0}^2 \leq \frac{1}{\lambda_2(\mathcal{L}_0)} \eta(x) \|x\|_{D_1}^2$. \square

The previous Lemma gives an upper bound on how “unbalanced” a good candidate for \mathbb{S}^{d-1} synchronization can be. Intuitively this should imply that, for most vertices i , x_i has roughly the same norm. What follows clarifies this thought.

Definition 7. Given $x \in \mathbb{R}^{dn}$, normalized so that $\|x\|_{D_1}^2 = \text{vol}(G)$, and a positive number δ , we define the Ill-balanced vertex subset of the graph G as $\mathcal{I}b_x(\delta) = \{i \in V : ||x_i|| - 1| \geq \delta\}$.

The volume of $\mathcal{I}b_x(\delta)$ is controlled by the following Lemma.

Lemma 8. Let $x \in \mathbb{R}^{dn}$ satisfy $\|x\|_{D_1}^2 = \text{vol}(G)$. Then,

$$\frac{\text{vol}(\mathcal{I}b_x(\delta))}{\text{vol}(G)} \leq \frac{4}{\delta^2} \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}.$$

Proof. Lemma 6 guarantees the existence of $\alpha_x \in \mathbb{R}_0^+$ such that $r_x = x - \alpha_x \tilde{x}$ satisfies $\|r_x\|_{D_1}^2 \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \|x\|_{D_1}^2$.

Let us start by bounding α_x ; by the triangle inequality,

$$(1 - \alpha_x)^2 \text{vol}(G) = (\|x\|_{D_1} - \alpha_x \|\tilde{x}\|_{D_1})^2 \leq \|r_x\|_{D_1}^2 \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \text{vol}(G),$$

which implies $(1 - \alpha_x)^2 \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}$.

If $i \in \mathcal{I}b_x(\delta)$ then $||x_i|| - 1| \geq \delta$, which implies $\|(r_x)_i\| = \|x_i\| - \alpha_x \geq ||x_i|| - 1| - |\alpha_x| \geq \delta - \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}}$. Squaring both sides of the inequality and summing over all $i \in \mathcal{I}b_x(\delta)$ gives,

$$\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \text{vol}(G) \geq \|r_x\|_{D_1}^2 \geq \sum_{i \in \mathcal{I}b_k} d_i \|(r_x)_i\|_2^2 \geq \text{vol}(\mathcal{I}b_k) \left(\delta - \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}} \right)^2, \quad (6)$$

as long as $\delta > \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}}$. Let us separate in two cases:

If $\frac{\delta}{2} > \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}}$, then, using (6) we have,

$$\frac{\text{vol}(\mathcal{I}b_k)}{\text{vol}(G)} \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \left(\delta - \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}} \right)^{-2} \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \left(\delta - \frac{\delta}{2} \right)^{-2} = \frac{4}{\delta^2} \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}.$$

If, on the other hand, $\frac{\delta}{2} \leq \sqrt{\frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}}$, then, since $\frac{\text{vol}(\mathcal{I}b_k)}{\text{vol}(G)} \leq 1$,

$$\frac{\text{vol}(\mathcal{I}b_k)}{\text{vol}(G)} \leq 1 \leq \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} \left(\frac{\delta}{2} \right)^{-2} = \frac{4}{\delta^2} \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}.$$

□

By having a bound on how large the set of Ill-balanced vertices can be (Lemma 8) we can control how much $\eta(x)$ is affected when we locally normalize x . This statement is made precise in the following Lemma, which finalizes the proof of the Theorem.

Lemma 9. For every $x \in \mathbb{R}^{dn}$, $\eta(\tilde{x}) \leq \frac{44}{\lambda_2(\mathcal{L}_0)} \eta(x)$.

Proof. We want to bound $\eta(\tilde{x}) = \frac{1}{2 \text{vol}(G)} \sum_{ij} w_{ij} \|\tilde{x}_i - \rho_{ij} \tilde{x}_j\|^2$. Without loss of generality we can assume $\|x\|_{D_1}^2 = \text{vol}(G)$. Let $0 < \gamma < 1$, then

$$\begin{aligned} \eta(\tilde{x}) &\leq \frac{1}{2 \text{vol}(G)} \left(\sum_{i \in \mathcal{I}b_x(\gamma)} \sum_j w_{ij} \|\tilde{x}_i - \rho_{ij} \tilde{x}_j\|^2 + \sum_{j \in \mathcal{I}b_x(\gamma)} \sum_i w_{ij} \|\tilde{x}_i - \rho_{ij} \tilde{x}_j\|^2 + \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \|\tilde{x}_i - \rho_{ij} \tilde{x}_j\|^2 \right) \\ &\leq 4 \frac{\text{vol}(\mathcal{I}b_x(\gamma))}{\text{vol}(G)} + \frac{1}{2 \text{vol}(G)} \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \|\tilde{x}_i - \rho_{ij} \tilde{x}_j\|^2. \end{aligned}$$

By Lemma 8 we have $4 \frac{\text{vol}(\mathcal{I}b_x(\gamma))}{\text{vol}(G)} \leq \frac{16}{\gamma^2} \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)}$.

Note that, for any $y, z \in \mathbb{R}^d$, $\left\| \frac{y}{\|y\|} - \frac{z}{\|z\|} \right\| \leq \frac{\|y-z\|}{\min\{\|y\|, \|z\|\}}$. By setting $y = x_i$ and $z = \rho_{ij}x_j$ we get $\|\tilde{x}_i - \rho_{ij}\tilde{x}_j\| \leq \frac{\|x_i - \rho_{ij}x_j\|}{\min\{\|x_i\|, \|x_j\|\}}$. This implies that

$$\begin{aligned} \frac{1}{2 \text{vol}(G)} \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \|\tilde{x}_i - \rho_{ij}\tilde{x}_j\|^2 &\leq \frac{1}{2 \text{vol}(G)} \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \left(\frac{\|x_i - \rho_{ij}x_j\|}{\min\{\|x_i\|, \|x_j\|\}} \right)^2 \\ &\leq \frac{1}{2 \text{vol}(G)} \frac{1}{(1-\gamma)^2} \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \|x_i - \rho_{ij}x_j\|^2. \end{aligned}$$

This means that

$$\begin{aligned} \eta(\tilde{x}) &\leq \frac{16}{\gamma^2} \frac{\eta(x)}{\lambda_2(\mathcal{L}_0)} + \frac{1}{2 \text{vol}(G)} \frac{1}{(1-\gamma)^2} \sum_{i,j \notin \mathcal{I}b_x(\gamma)} w_{ij} \|x_i - \rho_{ij}x_j\|^2 \\ &\leq \left(\frac{16}{\gamma^2} \frac{1}{\lambda_2(\mathcal{L}_0)} + \frac{1}{(1-\gamma)^2} \right) \eta(x). \end{aligned}$$

Since $\lambda_2(\mathcal{L}_0) \leq 1$ (see, e.g., [6]), it is possible to pick γ (e.g. 0.7) such that $\frac{16}{\gamma^2} \frac{1}{\lambda_2(\mathcal{L}_0)} + \frac{1}{(1-\gamma)^2} \leq \frac{44}{\lambda_2(\mathcal{L}_0)}$. This finalizes the proof. \square

Lemma 9 immediately implies the bound for η_G in Theorem 5. In order to obtain the bound for ζ_G one can simply note that, if using Lemma 3 for \tilde{x} , one has $\tilde{x}^t = \tilde{x}$, for every $t > 0$. Thus $\zeta(\tilde{x}) \leq \sqrt{\frac{5}{2}\eta(\tilde{x})}$.

3 The $O(d)$ Cheeger inequality

We present now our main result,

Theorem 10. *Let $\lambda_i(\mathcal{L}_1)$ and $\lambda_i(\mathcal{L}_0)$ denote the i -th smallest eigenvalue of, respectively, the normalized Connection Laplacian \mathcal{L}_1 and the normalized graph Laplacian \mathcal{L}_0 . Let ϑ_G and ν_G denote, respectively, the $O(d)$ frustration ℓ_1 constant and the $O(d)$ frustration ℓ_2 constant of G . Then,*

$$\frac{1}{d} \sum_{i=1}^d \lambda_i(\mathcal{L}_1) \leq \vartheta_G \leq 102d^2 \sqrt{\frac{1}{\lambda_2(\mathcal{L}_0)} \sum_{i=1}^d \lambda_i(\mathcal{L}_1)},$$

and

$$\frac{1}{d} \sum_{i=1}^d \lambda_i(\mathcal{L}_1) \leq \nu_G \leq 1026d^3 \frac{1}{\lambda_2(\mathcal{L}_0)} \sum_{i=1}^d \lambda_i(\mathcal{L}_1).$$

For the \mathbb{S}^{d-1} case, the main difficulty that we faced in trying to obtain, from eigenvectors, candidate solutions for the localization problem was the local unit norm constraint. This is due to the fact that the localization problem requires its solution to be a function from V to \mathbb{S}^{d-1} , corresponding to a vector in \mathbb{R}^{dn} whose vertex subvectors have unit norm, while the eigenvector, in general, does not satisfy such a constraint. Nevertheless, the results in the previous section show that, by simply rounding the eigenvector, one does not lose more than a linear term (if considering η_G), given that the graph has a spectral gap bounded away from zero.

However, the $O(d)$ synchronization setting is harder. The reason being that, besides the local normalization constraint, there is also a local orthogonality constraint (at each vertex, the d vectors have to be orthogonal

so that they can be the columns of an orthogonal matrix). For \mathbb{S}^{d-1} we locally normalized the vectors, by choosing for each vertex the unit vector closest to x_i . For $O(d)$ synchronization we will pick, for each vertex, the orthogonal matrix closest (in the Frobenius norm) to the matrix $[x_i^1 \cdots x_i^d]$. This can be achieved by the Polar decomposition. Given a $d \times d$ matrix X , the matrix $U(X)$, solution of $\min_{U \in O(d)} \|U(X) - X\|_F$, is one of the components of the Polar decomposition of X (see [9, 10] and references therein) but, slightly abusing notation, we will call it the Polar decomposition of X . We note that $U(X)$ can be computed efficiently through the SVD decomposition of X . In fact, given the SVD decomposition of X , $X = U\Sigma V^T$, the polar decomposition of X is given by $U(X) = UV^T$ (see [9]).

We will show the following Lemma

Lemma 11. *Given $x^1, \dots, x^d \in \mathbb{R}^{dn}$ such that $\langle x^k, x^l \rangle_{D_1} = 0$ for all $k \neq l$, consider the potential $O : V \rightarrow O(d)$ given as $O_i = U(X_i)$ where $X_i = [x_i^1 \cdots x_i^d]$ and $U(X)$ is the Polar decomposition of X , the orthogonal matrix that is closest to X in the Frobenius norm. If X_i is singular⁵ $U(X_i)$ is simply set to be I_d . Then,*

$$\nu(O) \leq (2d^{-1} + 2^{10}d^3) \frac{1}{\lambda_2(\mathcal{L}_0)} \sum_{i=1}^d \eta(x^i).$$

We need the following result of Li [10].

Lemma 12 (Theorem 1 in [10]). *Let $A, B \in \mathbb{C}^{d \times d}$ be non-singular matrices with polar decompositions $A = U(A)P$ and $B = U(B)P'$. Then $\|U(A) - U(B)\|_F \leq \frac{2}{\sigma_{\min}(A) + \sigma_{\min}(B)} \|A - B\|_F$, where $\sigma_{\min}(A)$ is the smallest singular value of the matrix A .*

In order to be able to bound from below the smallest singular value of $X = [x_i^1 \cdots x_i^d]$ we will introduce a notion similar to $\mathcal{I}b_x(\delta)$ but designed to take into account local orthogonality instead of local normalization.

Definition 13. *Given $x, y \in \mathbb{R}^{dn}$ two D_1 -orthogonal vectors, normalized so that $\|x\|_{D_1}^2 = \|y\|_{D_1}^2 = \text{vol}(G)$, and a positive number δ , we defined the Ill-balanced vertex subset of the graph G as*

$$\mathcal{I}b_{xy}(\delta) = \{i \in V : |\langle x_i, y_i \rangle| \geq \delta\}.$$

The following Lemma illustrates the purpose of this definition.

Lemma 14. *Let $x^1, \dots, x^d \in \mathbb{R}^{dn}$ be D_1 -orthogonal vectors, normalized so that $\|x^k\|_{D_1}^2 = \text{vol}(G)$. Let us define the “balanced” set \mathcal{B} as the complement of $\bigcup_{k \in [d]} \left(\mathcal{I}b_{x^k} \left(\frac{1}{8d} \right) \cup \bigcup_{m \in [d] \setminus \{k\}} \mathcal{I}b_{x^k x^m} \left(\frac{1}{2d} \right) \right)$. For all $i, j \in \mathcal{B}$, we have*

$$\|U(X_i) - \rho_{ij}U(X_j)\|_F \leq \sqrt{2}\|X_i - \rho_{ij}X_j\|_F.$$

Proof. For $i \in \mathcal{B}$, consider the gram matrix $X_i^T X_i$. Its k -th diagonal entry satisfies $\|x_i^k\|^2 \geq (1 - \frac{1}{8d})^2 \geq 1 - \frac{1}{4d}$. On the other hand the non-diagonal entries are, in magnitude, smaller or equal to $\frac{1}{2d}$. By the Gershgorin circle theorem, the smallest eigenvalue of $X_i^T X_i$, which is equal to $\sigma_{\min}(X_i)^2$, satisfies $\sigma_{\min}(X_i)^2 \geq 1 - \frac{1}{4d} - (d-1)\frac{1}{2d}$. Hence, $\sigma_{\min}(X_i) \geq \frac{1}{\sqrt{2}}$. By observing that $U(\rho_{ij}X_j) = \rho_{ij}U(X_j)$, and using Lemma 12, we get $\|U(X_i) - \rho_{ij}U(X_j)\|_F \leq \sqrt{2}\|X_i - \rho_{ij}X_j\|_F$. \square

Now we want to bound the size of $\mathcal{I}b_{xy}(\delta)$.

Lemma 15. *Let $x, y \in \mathbb{R}^{dn}$ be D_1 -orthogonal vectors such that $\|x\|_{D_1}^2 = \|y\|_{D_1}^2 = \text{vol}(G)$. Then,*

$$\frac{\text{vol}(\mathcal{I}b_{xy}(\frac{1}{2d}) \setminus (\mathcal{I}b_x(\frac{1}{8d}) \cup \mathcal{I}b_y(\frac{1}{8d})))}{\text{vol}(G)} \leq 4(8d)^2 \frac{\eta(x) + \eta(y)}{\lambda_2(\mathcal{L}_0)}.$$

⁵In this case the uniqueness of $U(X_i)$ is not guaranteed and thus the map is not well-defined.

Proof. Let us consider the vector $u = \frac{1}{\sqrt{2}}(x + y)$. It satisfies $\|u\|_{D_1}^2 = \text{vol}(G)$ and by the triangle inequality on the norm $\|\cdot\|_{L_1}$, $\eta(u) \leq \eta(x) + \eta(y)$. By Lemma 8 we get $\frac{\text{vol}(\mathcal{I}b_u(\frac{1}{8d}))}{\text{vol}(G)} \leq 4(8d)^2 \frac{\eta(x) + \eta(y)}{\lambda_2(\mathcal{L}_0)}$. We conclude the proof by noting that $\mathcal{I}b_{xy}(\frac{1}{2d}) \subset \mathcal{I}b_x(\frac{1}{8d}) \cup \mathcal{I}b_y(\frac{1}{8d}) \cup \mathcal{I}b_u(\frac{1}{8d})$. In fact, if $i \notin \mathcal{I}b_x(\frac{1}{8d}) \cup \mathcal{I}b_y(\frac{1}{8d}) \cup \mathcal{I}b_u(\frac{1}{8d})$ then $|\langle x_i, y_i \rangle| = \left| \|u_i\|^2 - \frac{\|x_i\|^2 + \|y_i\|^2}{2} \right| \leq \left(1 + \frac{1}{8d}\right)^2 - \left(1 - \frac{1}{8d}\right)^2 = \frac{1}{2d}$. \square

Proof. [of Lemma 11]

Let us consider \mathcal{B} as defined in Lemma 14, meaning $\mathcal{B}^c = \bigcup_{k \in [d]} \left(\mathcal{I}b_{x^k}(\frac{1}{8d}) \cup \bigcup_{m \in [d] \setminus \{k\}} \mathcal{I}b_{x^k x^m}(\frac{1}{2d}) \right)$. We want to bound $\nu(O) = \frac{1}{2d \text{vol}(G)} \sum_{ij} w_{ij} \|O_i - \rho_{ij} O_j\|_F^2$. Since $(\mathcal{B} \times \mathcal{B})^c \subset (\mathcal{B}^c \times V) \cup (V \times \mathcal{B}^c)$,

$$\begin{aligned} \sum_{ij} w_{ij} \|O_i - \rho_{ij} O_j\|_F^2 &\leq \sum_{(i,j) \in \mathcal{B} \times \mathcal{B}} w_{ij} \|U(X_i) - \rho_{ij} U(X_j)\|_F^2 + 2 \sum_{i \in \mathcal{B}^c} \sum_{j \in V} w_{ij} \|O_i - \rho_{ij} O_j\|_F^2 \\ &\leq 2 \sum_{ij} w_{ij} \|X_i - \rho_{ij} X_j\|_F^2 + 8d \text{vol}(\mathcal{B}^c), \end{aligned}$$

where the second inequality was obtained by using Lemma 14 and noting that O_i is an orthogonal matrix. To bound $\text{vol}(\mathcal{B}^c)$ we make use of Lemmas 8 and 15 and get that $\frac{\text{vol}(\mathcal{B}^c)}{\text{vol}(G)}$ is bounded above by

$$\sum_{k \in [d]} \frac{\text{vol}(\mathcal{I}b_{x^k}(\frac{1}{8d}))}{\text{vol}(G)} + \frac{1}{2} \sum_{k \in [d]} \sum_{m \in [d] \setminus \{k\}} \frac{\text{vol}(\mathcal{I}b_{x^k x^m}(\frac{1}{2d}) \setminus (\mathcal{I}b_x(\frac{1}{8d}) \cup \mathcal{I}b_y(\frac{1}{8d})))}{\text{vol}(G)} \leq 2^8 d^3 \frac{\sum_{k \in [d]} \eta(x^k)}{\lambda_2(\mathcal{L}_0)}.$$

Since $\sum_{ij} w_{ij} \|X_i - \rho_{ij} X_j\|_F^2 = 2 \text{vol}(G) \sum_{k=1}^d \eta(x^k)$, we get $\nu(O) \leq (2d^{-1} + 2^{10} d^3) \frac{1}{\lambda_2(\mathcal{L}_0)} \sum_{k=1}^d \eta(x^k)$. \square

Proof. [of Theorem 10]

The lower bound for ν_G is a simple consequence of the fact that the eigenvector problem is a relaxation of the Synchronization one and the upper bound is a direct consequence of Lemma 11.

The lower and upper bounds for ϑ_G can be obtained by comparing ϑ_G with $\nu_G(O)$. It is easy to check that $\vartheta_G(O) \geq \nu_G(O)$, which gives the lower bound.

On the other hand, denoting the k -th column of O_i by $(O_i)_{\cdot k}$, we have

$$\begin{aligned} \vartheta_G(O) &= \frac{1}{\sqrt{d} \text{vol}(G)} \sum_{ij} w_{ij} \sqrt{\sum_{k=1}^d \|(O_i)_{\cdot k} - \rho_{ij} (O_j)_{\cdot k}\|^2} \\ &\leq \frac{2}{\sqrt{d}} \sum_{k=1}^d \frac{1}{2 \text{vol}(G)} \sum_{ij} w_{ij} \|(O_i)_{\cdot k} - \rho_{ij} (O_j)_{\cdot k}\| \\ &\leq \frac{2}{\sqrt{d}} \sum_{k=1}^d \sqrt{\frac{5}{2} \frac{1}{2 \text{vol}(G)} \sum_{ij} w_{ij} \|(O_i)_{\cdot k} - \rho_{ij} (O_j)_{\cdot k}\|^2} \end{aligned} \tag{7}$$

$$\begin{aligned} &\leq \sqrt{10} \sqrt{\frac{d}{2d \text{vol}(G)} \sum_{k=1}^d \sum_{ij} w_{ij} \|(O_i)_{\cdot k} - \rho_{ij} (O_j)_{\cdot k}\|^2} \\ &= \sqrt{10} d^{1/2} \nu_G(O)^{1/2}, \end{aligned} \tag{8}$$

where (7) is obtained using Lemma 3 (via the same observation as was made in the proof of Theorem 5) and (8) is a direct consequence of Cauchy Schwartz inequality. This gives the upper bound for ϑ_G and thus completes the proof. \square

4 Spectral $O(d)$ Synchronization Algorithm

The proof of Theorem 10 suggests the following algorithm to solve $O(d)$ Synchronization.

Algorithm 16. *Given a weighted graph $G = (V, E)$ and a function $\rho : E \rightarrow O(d)$, build the graph Connection Laplacian L_1 and the degree matrix D_1 . Compute x^1, \dots, x^d as the first d (D_1 -orthogonal) minimizers of $\frac{x^T L_1 x}{x^T D_1 x}$ (that correspond to an eigenvector problem). Then, for each vertex i , compute the polar decomposition of the $d \times d$ matrix at that vertex, obtaining the orthogonal matrix O_i .*

The proof of Theorem 10 guarantees that, if the graph G is sufficiently connected and ν_G is small, then Algorithm 16 will compute a good solution.

We note that an $SO(3)$ version of this Algorithm was proposed in [12] (without performance guarantees) and also used in [18]. Also, a very similar $SO(2)$ version was introduced in [11]. In [11] the resilience to a certain random noise model is analyzed.

The performance guarantee for this algorithm relies on the fact that the vectors obtained are D_1 -orthogonal. However, in practice, due to possible errors in the calculations this condition might be perturbed and the D_1 inner products of these vectors, although small, may no longer be exactly zero. It is easy to adapt the analysis to this setting and show that it is in fact robust to such perturbations.

The results in this paper also suggest an alternative to Algorithm 16 which corresponds to, instead of solving the eigenvector problem, determine the d vectors sequentially and, at each step, constrain on the vector being locally orthogonal to the previous ones (this can still be done efficiently, see [8]). After the d vectors are obtained one can simply locally normalize each one and output that as the Synchronization solution candidate. The issue with this method is that its iterative nature ⁶ makes its analysis more difficult, as it is hard to guarantee that small errors in the first few vectors would not greatly affect the remaining ones. Also, numerical simulations suggest that the performance, in practice, of both methods is roughly the same.

5 Tightness of results

Let us consider the ring graph on n vertices $G_n = (V_n, E_n)$ with $V_n = [n]$ and $E = \{(i, (i+1) \bmod n), i \in [n]\}$ with the edge weights all equal to 1 and $\rho : V \rightarrow O(d)$ as $\rho_{(n,1)} = -I$ and $\rho = I$ for all other edges. Define $x \in \mathbb{R}^{dn}$ by $x_k = [2\frac{k}{n} - 1, 0, \dots, 0]^T$. It is easy to check that $\eta(x) = \mathcal{O}(n^{-2})$ and that, for any $t > 0$, if $x^t \neq 0$, there will have to be at least one edge that is not compatible with x^t , implying $\zeta(x^t) \geq \frac{1}{2n}$ and $\eta(x^t) \geq \frac{1}{2n}$. This shows that the $1/2$ exponent in Lemma 3 is needed. In fact, by adding a few more edges to the graph G_n one can also show the tightness of Theorem 2: Consider the “rainbow” graph H_n that is constructed by adding to G_n , for each non-negative integer k smaller than $n/2$, an edge between vertex k and vertex $n - k$ with $\rho_{(k,n-k)} = -I$. The vector x still satisfies $\eta(x) = \mathcal{O}(n^{-2})$, however, for any non-zero vector $v : V \rightarrow \mathbb{S}^{d-1} \cup \{0\}$, it is not hard to show that both $\zeta(v)$ and $\eta(v)$ have to be of order at least n^{-1} , meaning that both ζ_G^* and η_G^* are $\Omega(\sqrt{\lambda_1(\mathcal{L}_1)})$. This also means that, even if considering η_G^* , one could not get a linear bound (like Lemma 9 provides) without the control on $\lambda_2(\mathcal{L}_0)$.

Theorem 10 provides a non-trivial bound only if $\lambda_2(\mathcal{L}_0)$ is sufficiently large. It is clear that if one wants to bound full frustration constants, a dependency on $\lambda_2(\mathcal{L}_0)$ is needed. It is, nevertheless, non-obvious that even if considering partial versions of $O(d)$ frustration constants, ϑ_G^* or ν_G^* , the dependency on $\lambda_2(\mathcal{L}_0)$ is still required. This can, however, be illustrated by a simple example in $O(2)$; consider a disconnected graph G with two sufficiently large complete components, $G^1 = (V^1, E^1)$ and $G^2 = (V^2, E^2)$. For each edge let $\rho_{i,j} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. It is clear that the vectors x^1 and x^2 defined such that $x_i^1 = [0, 1_{V^1(i)}]^T$ and $x_i^2 = [0, 1_{V^2(i)}]^T$ are orthogonal to each other and lie in the null space of the graph Connection Laplacian

⁶The fact that the calculation of one of the vectors greatly depends on the ones already computed.

of G . This implies that $\lambda_2(\mathcal{L}_1) = 0$. On the other hand, it is straightforward to check that neither ϑ_G^* or ν_G^* are zero because it is impossible to perfectly synchronize the graph (or even any of the components).

6 Concluding Remarks

Synchronization is a challenging problem. Recent discoveries [12] suggest that spectral methods are very promising as a feasible method to solve this problem. In fact, in [11], probability guarantees of performance are given for the performance of a spectral method to solve the $SO(2)$ synchronization problem under a certain random noise model. Nevertheless, to the best of our knowledge, Algorithm 16 is the first method for $O(d)$ synchronization having a (deterministic) worst case performance guarantee. As one would expect, the worst case performance is significantly weaker than the kind of probabilistic guarantees, given a specific noise model, e.g. as the one given in [11]. In fact, the guarantee in [11] is given in terms of distance between the candidate solution and the ground truth, while the one on Algorithm 16 is given in terms of the compatibility error.

In special applications one knows, a priori, that every element in the potential has positive determinant. This corresponds to a synchronization problem in $SO(d)$. Although this can be viewed as a special case of the $O(d)$ problem it is expected that the additional structure can be leveraged to improve the algorithm (and the analysis). In particular, the first $d - 1$ columns of a matrix in $SO(d)$ completely determine the matrix. This suggests that the $SO(d)$ synchronization problem is solvable by just the first $d - 1$ eigenvectors of the Connection Laplacian, instead of the d first ones. In fact, the $SO(2)$ Synchronization problem is equivalent to the \mathbb{S}^1 localization one, and the guarantees for the \mathbb{S}^1 localization problem were given solely in terms of the first eigenvalue of the graph Connection Laplacian. We leave the improved $SO(d)$ analysis for future work.

One might argue that, in some applications, the weights on the edges of the graph do not have a clear meaning. The reason being that we may be given a few relative measurements ρ_{ij} and it is unclear how to give weights to such measurements. In such cases, since we have the freedom of choosing the weights of the edges and Theorem 16 suggests that our method will work better with a large $\lambda_2(\mathcal{L}_0)$, one could compute the weights of the edges in a way such that $\lambda_2(\mathcal{L}_0)$ is maximized. This problem is solved in [16]. The caveat is that, the new weights will affect the way the compatibility error is measured and the eigenvalues of the Connection Laplacian. It is thus still unclear if such an approach would improve the method. Another interesting possible outcome of a procedure of this nature is a possible ranking of the edges, large weights would likely tend to be given to edges that are more important to ensure the connectivity of the graph.

The classical Cheeger Inequality has an analogous result in smooth manifolds (actually, the first to be shown [4]). One interesting question is whether the Theorems in this paper have an analogous smooth version. One difficulty is to understand what would correspond to the frustration constant on the smooth case. Vector Diffusion Maps [13] suggest that an analogous result in smooth manifold would be related to the parallel transport and its incompatibility and some results in Differential Geometry [3] suggest that the Holonomy could be a geometric property that corresponds to the frustration constant. These suggestions are “coherent” because Holonomy can, in some sense, be viewed as the incompatibility of the Parallel transport (due to the curvature of the manifold).

Acknowledgments

The authors thank Hau-Tieng Wu for interesting discussions on the topic of this paper. Afonso S. Bandeira was supported by Award Number DMS-0914892 from the NSF. Amit Singer was partially supported by Award Number DMS-0914892 from the NSF, by Award Number FA9550-09-1-0551 from AFOSR, by Award Number R01GM090200 from the National Institute of General Medical Sciences, and by the Alfred P. Sloan Foundation. Daniel A. Spielman was supported by the NSF under Grant No. 091548.

References

- [1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.
- [2] N. Alon and V. Milman. Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory*, 38:73–88, 1985.
- [3] W. Ballmann, J. Brüning, and G. Carron. Eigenvalues and holonomy. *Int. Math. Res. Not.*, (12):657–665, 2003.
- [4] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pp. 195–199. Princeton Univ. Press, 1970.
- [5] F. Chung. Four proofs for the cheeger inequality and graph partition algorithms. *Fourth International Congress of Chinese Mathematicians*, pp. 331–349, 2010.
- [6] F. Chung and L. Lu. *Complex graphs and networks*. Number 107. Amer Mathematical Society, 2006.
- [7] M. Cucuringu, Y. Lipman, and A. Singer. Sensor network localization by eigenvector synchronization over the euclidean group. *ACM Transactions on Sensor Networks*, In press.
- [8] G. H. Golub. Some modified matrix eigenvalue problems. *SIAM Rev.*, 15(2), 1973.
- [9] N. J. Higham. Computing the polar decomposition with applications. *SIAM J. Sci. Stat. Comput.*, 7:1160–1174, October 1986.
- [10] R.-C. Li. New perturbation bounds for the unitary polar factor. *SIAM J. Matrix Anal. Appl.*, 16(1):327–332, January 1995.
- [11] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, 30(1):20 – 36, 2011.
- [12] A. Singer and Y. Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM J. Imaging Sciences*, 4(2):543–572, 2011.
- [13] A. Singer and H.-T. Wu. Vector diffusion maps and the connection laplacian. *Comm. Pure Appl. Math.*, in press.
- [14] A. Singer and H.-T. Wu. Orientability and diffusion maps. *Appl. Comput. Harmon. Anal.*, 31(1):44–58, 2011.
- [15] A. Singer, Z. Zhao, , Y. Shkolnisky, and R. Hadani. Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM Journal on Imaging Sciences*, 4:723–759, 2011.
- [16] J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Review*, 48(4):681–699, 2006.
- [17] L. Trevisan. Max cut and the smallest eigenvalue. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC ’09, pages 263–272, New York, NY, USA, 2009. ACM.
- [18] T. Tzeneva. Global alignment of multiple 3-d scans using eigenvector synchronization. *Senior Thesis, Princeton University (supervised by S. Rusinkiewicz and A. Singer)*, 2011.